Primal robustness and semidefinite cones

Seungil You, Ather Gattami, and John C. Doyle

***Abstract*— This paper reformulates and streamlines the core tools of robust stability and performance for LTI systems using now-standard methods in convex optimization. In particular, robustness analysis can be formulated directly as a primal convex (semidefinite program or SDP) optimization problem using sets of gramians whose closure is a semidefinite cone. This allows various constraints such as structured uncertainty to be included directly, and worst-case disturbances and perturbations constructed directly from the primal variables. Well known results such as the KYP lemma and various scaled small gain tests can also be obtained directly through standard SDP duality. To readers familiar with robustness and SDPs, the framework should appear obvious, if only in retrospect. But this is also part of its appeal and should enhance pedagogy, and we hope suggest new research. There is a key lemma proving closure of a grammian that is also obvious but our current proof appears unnecessarily cumbersome, and a final aim of this paper is to enlist the help of experts in robust control and convex optimization in finding simpler alternatives.**

I. Introduction

Robust control theory has been an important subject in the control engineering both in theory and practice [1]. Theoretical developments has been evolved in various flavor, but the most modern approach to this subject is the Linear Matrix Inequality (LMI) based approach [2].

In order to obtain the LMI characterization of system behavior, so called the $\mathcal{S}$-procedure [3] has been extensively used. Motivated by the popularity of LMIs in systems theory, semidefinite programming (SDP) duality has been used to understand such LMIs and control theoretic interpretation of the SDP duals of these LMIS has been reported [4], [5], [6]. In particular, the dual LMI approach is used to extract the worst case frequency variable and disturbance in [5], [6]. However, a recent paper shows that this dual LMI has its own right as a well-defined optimization when it comes to $\mathcal{H}_\infty$ analysis [7]. In [8], we also report that the Kalman–Yakubovich–Popov (KYP) lemma [9] is an SDP dual of this optimization problem, which should be obvious to the experts. An interesting observation in here is that the dual LMI may be another starting point of robustness analysis which does not require $\mathcal{S}$-procedure, and well known results, such as the KYP lemma can be obtained through SDP duality, *i.e.*, reversing the theoretical developments.

To this end, this paper provides a complete characterization of gramians generated by a linear time invariant (LTI) system. It turns out the closure of a set of gramians is an intersection of a subspace and a semidefinite cone. The seminal paper [10] attempts to obtain similar results on the covariance matrices generated by stochastic disturbances, but this paper characterizes gramians from *deterministic* disturbances, which is suitable for existing robustness results. More importantly, we provide a semidefinite representation of gramians in contrast to the rank constraint in [10]. This semidefinite representation allows us to formulate extended $\mathcal{H}_\infty$ analysis, where we can directly capture numerous prior information on the disturbance including structural properties, as an SDP.

In addition, the SDP dual of our primal optimization provides the well known LMI characterization of the system behavior. We exemplify this procedure for the KYP lemma, but the result can be easily extended to more general disturbance setting, and our approach provides an arguably simple proof through standard SDP duality. In addition, SDP duality also provides the scaled small gain tests for the robust stability verification. However, our primal formulation provides a specific input-output pair proving that the system is not robustly stable, which is not a trivial task in the scaled small gain test. This is because the variables in the scaled small gain test does not contain useful input-output information, so if the test fails, it is hard to extract a specific pair that disproves robust stability. This entire procedure of obtaining LMIs for robustness analysis should be obvious to the experts, which shows a pedagogical benefit of our approach. We also hope that our new tool opens up a new research direction in robust control theory.

A. Notation

$\mathbb{H}^n, \mathbb{H}^n_+, \mathbb{H}^n_{++}$ are sets of $n \times n$ Hermitian, positive semidefinite, positive definite matrices, respectively. The generalized inequality $X \succeq 0$ means $X \in \mathbb{H}_+$, and $X \succ 0$ means $X \in \mathbb{H}_{++}$. We use l_2 for $l_2[0,\infty)$, the Hilbert space of square summable sequence with the starting index 0. The bold Latin letter $\mathbf{x}$ means a sequence in l_2. In addition, for a vector and vector-valued signal, $\|\cdot\|_2$ is the two norm. For a matrix and linear operator, $\|\cdot\|_F$ is the Frobenius norm, and $\rho(A)$ denotes the spectral radius of A, A^* denotes a Hermitian/Adjoint operator.

II. A semidefinite representation of Gramians

For a signal $\mathbf{u} \in l_2$, we define the gramian $\Lambda : l_2 \to \mathbb{H}_+$,

$$\Lambda(\mathbf{u}) = \sum_{k=0}^{\infty} u_k u_k^*.$$

Notice that the gramian is well-defined because each entry of the matrix is finite, and $\mathbb{H}_+$ is closed.

This work was supported by Kwanjeong Graduate Fellowship

S. You and J. Doyle are with the Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA {syou, doyle}@caltech.edu A. Gattami is with Ericsson Research, Stockholm, Sweden, ather.gattami@ericsson.com

For notational convenience, let $\Lambda(\mathbf{u}, \mathbf{v}) = \Lambda\left(\begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}\right) = \sum_{k=0}^{\infty} \begin{bmatrix}u_k\\ v_k\end{bmatrix}\begin{bmatrix}u_k\\ v_k\end{bmatrix}^*$, and $\Lambda_N(\mathbf{u})$ be a finite truncation of Λ, $\Lambda_N(\mathbf{u}) = \sum_{k=0}^{N-1} u_k u_k^*$.

Suppose we have matrices $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, where A is Schur stable, $\rho(A) < 1$, and let $\mathbf{M} : l_2^m \to l_2^n$ be a linear operator such that $\mathbf{x} = \mathbf{M}\mathbf{w}$ if

$$x_{k+1} = Ax_k + Bw_k \tag{1}$$

$$x_0 = 0. \tag{2}$$

In this paper, we consider a set of gramians, $\mathcal{S}$, generated by $\mathbf{M}$,

$$\mathcal{S} := \{V \in \mathbb{H}_+ \ : V = \Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) \text{ for some } \mathbf{w} \in l_2\}.$$

In terms of $\mathbf{x}, \mathbf{w}$, the gramian can be seen as

$$V = \begin{bmatrix}X & R\\ R^* & W\end{bmatrix} = \begin{bmatrix}\sum_k x_k x_k^* & \sum_k x_k w_k^*\\ \sum_k w_k x_k^* & \sum_k w_k w_k^*\end{bmatrix}.$$

From this definition, we can easily see that the gramian captures various input-output relationships in the system. For example, $\|\mathbf{w}\|_2^2 = \mathbf{Tr}(W)$, and $\|C\mathbf{x} + D\mathbf{w}\|_2^2 = \mathbf{Tr}\left(\begin{bmatrix}C^*C & C^*D\\ D^*C & D^*D\end{bmatrix} V\right)$.

However, for a given matrix V, checking $V \in \mathcal{S}$ is not a trivial task, because one should search over the l_2 space, an infinite dimensional space, to find a signal $\mathbf{w}$ generating V. Therefore it is desirable to find a convinient way to characterize the set $\mathcal{S}$. Let us consider the following set

$$\mathcal{C} := \left\{V \in \mathbb{H}_+ : \begin{bmatrix}A & B\end{bmatrix} V \begin{bmatrix}A^*\\ B^*\end{bmatrix} = \begin{bmatrix}I & 0\end{bmatrix} V \begin{bmatrix}I\\ 0\end{bmatrix}\right\}.$$

Notice that $\mathcal{C}$ is an intersection of a subspace in $\mathbb{H}$ and a semidefinite cone, therefore $\mathcal{C}$ is a finite dimensional closed, convex cone that is semidefinite programming (SDP) representable. This means that checking $V \in \mathcal{C}$ can be easily done by a semidefinite programming.

Why do we need this set $\mathcal{C}$? Is there any relationship between $\mathcal{C}$ and $\mathcal{S}$? The following proposition shows an interesting observation between these two sets.

Proposition 1: *The set $\mathcal{S}$ is a subset of $\mathcal{C}$.*

Proof: Suppose $V \in \mathcal{S}$. This means there exists a signal $\mathbf{w} \in l_2$ such that $V = \Lambda(\mathbf{M}\mathbf{w}, \mathbf{w})$. Let $\mathbf{x} = \mathbf{M}\mathbf{w}$. Then, since $x_{k+1} = Ax_k + Bw_k$, we have

$$x_{k+1}x_{k+1}^* = \begin{bmatrix}A & B\end{bmatrix}\begin{bmatrix}x_k\\ w_k\end{bmatrix}\begin{bmatrix}x_k\\ w_k\end{bmatrix}^*\begin{bmatrix}A^*\\ B^*\end{bmatrix}$$

By taking an infinite sum, we have

$$\begin{aligned}\sum_{k=0}^{\infty} x_{k+1}x_{k+1}^* &= \sum_{k=0}^{\infty}\begin{bmatrix}A & B\end{bmatrix}\begin{bmatrix}x_k\\ w_k\end{bmatrix}\begin{bmatrix}x_k\\ w_k\end{bmatrix}^*\begin{bmatrix}A^*\\ B^*\end{bmatrix}\\ &= \begin{bmatrix}A & B\end{bmatrix} V \begin{bmatrix}A^*\\ B^*\end{bmatrix}.\end{aligned}$$

Moreover, since $x_0 = 0$, $\sum_{k=0}^{\infty} x_{k+1}x_{k+1}^* = \sum_{k=0}^{\infty} x_k x_k^*$. Using this fact with $\begin{bmatrix}I_n & 0_{n,m}\end{bmatrix} V \begin{bmatrix}I_n & 0_{n,m}\end{bmatrix}^* = \sum_{k=0}^{\infty} x_k x_k^*$, we can conclude that $V \in \mathcal{C}$. ∎

The above observation is somewhat trivial due to the convergence. But the immediate, and important question arises. Is $\mathcal{C}$ equal to $\mathcal{S}$? If this is true, we can replace a complicated set $\mathcal{S}$ by a semidefinite representable convex cone $\mathcal{C}$. For an optimization with $\mathcal{S}$, this has a dramatic impact: any optimization involving $\mathcal{S}$ becomes a convex program. Unfortunately, this is not the case.

Example 1: *Let $A = a$, $B = 1$, where $|a| < 1$, and $a \neq 0$. Consider $V = \begin{bmatrix}\frac{1}{1-a}\\ 1\end{bmatrix}\begin{bmatrix}\frac{1}{1-a}\\ 1\end{bmatrix}^*$. It can be easily checked $V \in \mathcal{C}$, and $\mathbf{Rank}(V) = 1$, but $V \notin \mathcal{S}$.*

However, we have a surprising fact: $\mathcal{C}$ is the *closure* of $\mathcal{S}$. Although the idea of the proof is simple, but our current proof goes through a tedious analysis to apply the $\epsilon - \delta$ style argument. So we present the main result here, and relegate the sketch of the proof to the appendix. But we want to emphasize that our proof is *constructive*, therefore given V, we can find a signal $\mathbf{w}$ that approximates V arbitrarily close.

Lemma 1: *Suppose $V = \begin{bmatrix}X & R\\ R^* & W\end{bmatrix} \in \mathcal{C}$. Then for all $\varepsilon > 0$, there exists $\mathbf{w} \in l_2$ with finite number of non-zero entries such that*

$$\|\Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) - V\|_F < \varepsilon \tag{3}$$

$$\Lambda(\mathbf{w}) = W \tag{4}$$

The following is immediate.

Lemma 2: $\mathcal{C} = \mathbf{cl}\,\mathcal{S}$.

Proof: From Proposition 1, $\mathcal{S} \subset \mathcal{C}$, which shows $\mathbf{cl}\,\mathcal{S} \subset \mathcal{C}$ since $\mathcal{C}$ is closed. From Lemma 1, $\mathcal{C} \subset \mathbf{cl}\,\mathcal{S}$. Therefore, $\mathcal{C} = \mathbf{cl}\,\mathcal{S}$. ∎

The above two results are key lemmas in this paper. Many robustness analysis can be stated as an optimization problem with gramians, that has a linear objective function. Therefore, a set of gramians, $\mathcal{S}$, can be replaced by $\mathcal{C}$ without any conservatism, and more importantly the resulting problem becomes an SDP.

Before concluding this section, we present a connection between the controllability of (A, B) and the geometric property of the SDP cone $\mathcal{C}$.

Proposition 2: *There exists $V \in \mathcal{C} \cap \mathbb{H}_{++}$ if and only if (A, B) is controllable.*

Proof: Suppose (A, B) is controllable. Since A is stable, the controllability gramian W_c

$$AW_cA^* - W_c + BB^* = 0$$

is positive definite. Let

$$V = \begin{bmatrix}W_c & 0\\ 0 & I\end{bmatrix},$$

then $V \in \mathcal{C} \cap \mathbb{H}_{++}$.

Now suppose there exists $V \in \mathcal{C} \cap \mathbb{H}_{+}$. Since $V \in \mathcal{C}$,

$$AXA^* - X + BR^*A^* + ARB^* + BWB^* = 0.$$

Let $W = TT^*$, and $\tilde{B} = BT$, $K = T^{-1}R^*X^{-1}$. Then,

$$(A + \tilde{B}K)X(A + \tilde{B}K)^* - X + \tilde{B}\tilde{B}^* = 0.$$

Since $X \succ 0$, $(A + \tilde{B}K, \tilde{B})$ is controllable, and it is easy to check that (A, B) is controllable. ∎

Now we use all these results to prove well-known results in robust control theory which shows the effectiveness of our new, primal approach.

III. $\mathcal{H}_\infty$ ANALYSIS AND THE KYP LEMMA

A. $\mathcal{H}_\infty$ analysis

In $\mathcal{H}_\infty$ analysis, we would like to find the worst-case disturbance that maximizes the output norm. Specifically, let $z_k = Cx_k + Dw_k$. Then we want to solve

$$\begin{aligned} \mu_\infty := &\underset{\mathbf{w},\mathbf{x},\mathbf{z}}{\text{maximize}} && \|\mathbf{z}\|_2^2 \\ &\text{subject to} && x_{k+1} = Ax_k + Bw_k, \qquad x_0 = 0 \\ & && \|\mathbf{w}\|_2^2 = 1. \end{aligned} \tag{5}$$

The optimal value of (5) is the square of the $\mathcal{H}_\infty$ norm of the system.

Let $\Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) = \begin{bmatrix} X & R \\ R^* & W \end{bmatrix}$, where $X \in \mathbb{H}^n$, $W \in \mathbb{H}^m$. Notice that

$$\begin{aligned} \|\mathbf{z}\|_2^2 &= \mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \right) \\ \|\mathbf{w}\|_2^2 &= \mathbf{Tr}\,(W)\,. \end{aligned}$$

This shows that the optimization (5) is equivalent to

$$\begin{aligned} &\underset{X,R,W}{\text{maximize}} && \mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \right) \\ &\text{subject to} && \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \in \mathcal{S} \\ & && \mathbf{Tr}\,(W) = 1. \end{aligned} \tag{6}$$

However the above problem is hard to solve because the feasible set $\mathcal{S}$ is involved with an infinite dimensional, l_2 space. Using Lemma 1 and 2, we can replace $\mathcal{S}$ by $\mathcal{C}$ which results in an SDP that computes the $\mathcal{H}_\infty$ norm of the system.

Proposition 3: *Define the set* $\mathcal{F} = \{V = \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \in \mathbb{H} : \mathbf{Tr}\,(W) = 1\}$. *Then* $\mathbf{cl}(\mathcal{S} \cap \mathcal{F}) = \mathcal{C} \cap \mathcal{F}$.

Proof: Since $\mathbf{cl}\,\mathcal{S} = \mathcal{C}$, and $\mathbf{cl}\,\mathcal{F} = \mathcal{F}$, we have $\mathbf{cl}\,(\mathcal{S} \cap \mathcal{F}) \subset \mathcal{C} \cap \mathcal{F}$. Now suppose $V \in \mathcal{C} \cap \mathcal{F}$. From Lemma 1, for any $\epsilon > 0$, there exists $\mathbf{w} \in l_2$ such that $\Lambda(\mathbf{w}) = 1$, and $\|V - \Lambda(\mathbf{M}\mathbf{w}, \mathbf{w})\|_F < \epsilon$. Since $\Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) \in \mathcal{S} \cap \mathcal{F}$, we can conclude that $V \in \mathbf{cl}(\mathcal{S} \cap \mathcal{F})$. ∎

Therefore, from the continuity of $\mathbf{Tr}\,(\cdot)$, we can replace $\mathcal{S}$ in (6) by $\mathcal{C}$ without any conservatism, and this is the main reason why we can compute the $\mathcal{H}_\infty$ norm using an SDP.

$$\begin{aligned} &\underset{X,R,W}{\text{maximize}} && \mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \right) \\ &\text{subject to} && X = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \begin{bmatrix} A^* \\ B^* \end{bmatrix} \\ & && \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \succeq 0, \mathbf{Tr}\,(W) = 1. \end{aligned} \tag{7}$$

Clearly, the above optimization is an SDP that can be solved via a standard SDP solver [11], and has the same form as [7], [8]. More importantly, without going through an additional proof, we can easily check the equivalence between the finite-dimensional SDP (7) and an infinite-dimensional optimization (6). Notice that after obtaining the optimal solution of (7), we can construct a signal $\mathbf{w}$ that asymptotically achieves the optimal value using the proof of Lemma 1. This direct formulation approach will be repeated through this paper, and our Lemma 1 provides a elementary, yet elegant approach to the classical problems in robust control theory.

Using SDP duality, we can expand our understanding of (7). The following is the SDP dual of (7):

$$\begin{aligned} &\underset{\lambda \in \mathbb{R}, P \in \mathbb{H}}{\text{minimize}} && \lambda \\ &\text{subject to} && \begin{bmatrix} A & B \\ C & D \end{bmatrix}^* \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} - \lambda \begin{bmatrix} 0 & 0 \\ 0 & I_m \end{bmatrix} \preceq 0. \end{aligned} \tag{8}$$

which is the same problem derived from the KYP lemma. Since we provide a well-defined optimization using our set $\mathcal{C}$, we can claim that the KYP lemma is the dual of $\mathcal{H}_\infty$ analysis. In addition, we can easily check that strong duality holds between (7) and (8) because the dual program (8) is strictly feasible.

Proposition 4: (8) *is strictly feasible.*

Proof: Since A is stable, there exists P such that $A^*PA - P + C^*C \prec 0$. Then by taking λ large enough, we can find a strictly feasible point of (8). ∎

As a result, we have the following corollary from the Conic duality theorem [12].

Corollary 1: *The duality gap between* (7), (8) *is zero, and the primal problem* (7) *is solvable.*

However the dual optimum may not be attained. Let us see the following example.

Example 2: *Let* $A = \frac{1}{2}, B = 0, C = 1, D = 1$. *Then the optimal solution of* (7) *is given by* $V^\star = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, *and the corresponding optimal value is* $+1$. *The optimal value of the dual* (8) *is also* $+1$, *by taking* $\lambda^\star = 1$, *and* $P^\star \to \infty$. *Clearly, the dual optimum is not attained.*

The pair (A, B) in the above example is not controllable, and this phenomena is closely related to the controllability assumption in the KYP lemma. In order to ensure the existence of a multiplier P (a dual optimal solution), we need the controllability assumption.

Proposition 5: *The primal program* (7) *is strictly feasible if and only if* (A, B) *is controllable.*

Proof: From Proposition 2, there exists $V \in \mathcal{C} \in \mathbb{H}_+$ if and only if (A, B) is controllable. ∎

As a corollary, we have the following result on strong duality.

Corollary 2: *Suppose (A,B) is controllable. Then both the primal problem* (7)*, and dual* (8) *are solvable.*

B. A proof of KYP lemma

Based on the observation between the KYP lemma and our primal optimization, it is very easy to prove the KYP lemma using the theorem of alternatives for SDP [4].

Let $\mathcal{A}$, $\mathcal{B}$ linear operators on $\mathbb{H}$. Then,

Theorem 1 (ALT4): *Exactly one of the following is true.*

(i) There exists an $X \in \mathbb{H}$ with $\mathcal{A}(X) + A_0 \succ 0$, and $\mathcal{B}(X) = 0$.

(ii) There exists a non zero $Z \in \mathbb{H}_+$, $W \in \mathbb{H}$, $\mathcal{A}^*(Z) + \mathcal{B}^*(W) = 0$, and $\mathbf{Tr}\,(A_0 Z) \leq 0$.

Now we are ready to prove the following theorem.

Theorem 2 (KYP lemma, strict inequality): *The optimal value of* (5)*, $\mu_\infty < 1$ if and only if there exists $P \in \mathbb{H}^n$ such that*

$$\begin{bmatrix} A^*PA - P & A^*PB \\ B^*PA & B^*PB \end{bmatrix} + \begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D - I_m \end{bmatrix} \prec 0. \quad (9)$$

Proof: From the Theorem 1, there exists P with (9) holds if and only if there is no non-zero $V \succeq 0$ such that $\begin{bmatrix} A & B \end{bmatrix} V \begin{bmatrix} A & B \end{bmatrix}^* = \begin{bmatrix} I & 0 \end{bmatrix} V \begin{bmatrix} I & 0 \end{bmatrix}^*$, and $\mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D - I \end{bmatrix} V \right) \geq 0$. Notice that this is equivalent to the optimal value of (7) is greater than equal to 1, since the optimum is always attained. Since (7) is equivalent to (5), we can conclude the proof. ■

Theorem 3 (KYP lemma, non-strict inequality): *Suppose (A,B) is controllable. Then $\mu_\infty \leq 1$ if and only if there exists $P \in \mathbb{H}^n$ such that*

$$\begin{bmatrix} A^*PA - P & A^*PB \\ B^*PA & B^*PB \end{bmatrix} + \begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D - I_m \end{bmatrix} \preceq 0. \quad (10)$$

Proof: If $\mu_\infty < 1$, then from the Theorem 3 the result is obvious. Suppose the optimal value of (5) is 1. Then the optimal value of (7) is also 1. From the Corollary 2, this is equivalent to the existence of the dual optimal solution $(\lambda, P) = (1, P^\star)$, and this concludes the proof. ■

IV. Extended $\mathcal{H}_\infty$ analysis

In $\mathcal{H}_\infty$ analysis, a disturbance $\mathbf{w}$ is assumed to have an unit energy, $\|\mathbf{w}\|_2 = 1$. Suppose more information about a disturbance is known beforehand. Then $\mathcal{H}_\infty$ norm becomes conservative since the analysis does not exploit this additional information. Therefore, it is natural to ask a question whether we can capture more general disturbance sets beyond $\|\mathbf{w}\|_2 = 1$, and formulate appropriate $\mathcal{H}_\infty$ optimization. In this section, we propose extended $\mathcal{H}_\infty$ analysis with different constraints on the disturbance. Some of these results are known from [13] in the form of the scaled small gain test, but we explicitly propose a well-defined optimization problem which contains a disturbance information which allows us to extract the worst-case disturbance, and provide its exactness without additional effort.

Now consider the following disturbance set.

$$\mathcal{W} = \{\mathbf{w} \in l_2 : f_i\left(\Lambda(\mathbf{w})\right) \preceq 0, i = 1, \cdots, n_c\},$$

where each f_i is a matrix valued linear function maps $\mathbb{H}$ to $\mathbb{H}$. This set can be used to capture some interesting prior knowledge on the disturbance.

Example 3 ($\boldsymbol{\mathcal{H}_\infty}$ analysis): *In the $\mathcal{H}_\infty$ analysis, we require $\|\mathbf{w}\|_2 = 1$. Using $f_1(W) = \mathbf{Tr}\,(W) - 1$, $f_2(W) = 1 - \mathbf{Tr}\,(W)$, then,*

$$\mathcal{W} = \left\{\mathbf{w} \in l_2 : \mathbf{Tr}\left(\Lambda(\mathbf{w})\right) = \|\mathbf{w}\|_2^2 = 1\right\},$$

which is desired.

Example 4 (Square $\boldsymbol{\mathcal{H}_\infty}$ analysis): *In [13], the disturbance $\mathbf{w}$ satisfies $\|\mathbf{w}_i\|_2 \leq 1$, for $i = 1, \cdots, m$, where $\mathbf{w}_i$ is the ith component of $\mathbf{w}$. Using $f_i(W) = W_{ii} - 1$, for $i = 1, \cdots, m$, then,*

$$\mathcal{W} = \left\{\mathbf{w} \in l_2 : f_i\left(\Lambda(\mathbf{w})\right) = \|\mathbf{w}_i\|_2^2 \leq 1, i = 1, \cdots, m\right\},$$

which is desired.

Example 5 (Grouped square $\boldsymbol{\mathcal{H}_\infty}$ analysis): *Suppose the disturbance is in l_2^4, and $\|\mathbf{w}_1\|_2^2 + \|\mathbf{w}_2\|_2^2 \leq 1$, $\|\mathbf{w}_3\|_2^2 + \|\mathbf{w}_4\|_2^2 \leq 1$. Then using $f_1(W) = W_{11} + W_{22} - 1$, $f_2(W) = W_{33} + W_{44} - 1$, we can capture this disturbance.*

Example 6 (Principal component bound): *Suppose we know that the maximum eigenvalue of the autocovariance of the signal $\mathbf{w}$ is bounded by one. This can be easily captured by $f_1(W) = W - I_m$. and $\mathcal{W} = \{\mathbf{w} \in l_2 : \Lambda(\mathbf{w}) \preceq I_m\}$. See [14] and references therein for the application of this disturbance modeling.*

These examples show that our modeling framework can capture various information on the disturbance. Now the next step is to find a computationally tractable method to analyze the worst-case performance as in the $\mathcal{H}_\infty$ case. In other words, we would like to find a way to solve the following infinite dimensional optimization.

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{x}}{\text{maximize}} & \|C\mathbf{x} + D\mathbf{w}\|_2^2 \\ \text{subject to} & x_{k+1} = Ax_k + Bw_k, x_0 = 0 \\ & \mathbf{w} \in \mathcal{W}. \end{array} \quad (11)$$

Using the Gramian $V = \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} = \Lambda(\mathbf{M}\mathbf{w}, \mathbf{w})$ as in the $\mathcal{H}_\infty$ analysis, the above optimization is equivalent to

$$\begin{array}{ll} \underset{X,R,W}{\text{maximize}} & \mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \right) \\ \text{subject to} & V = \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \in \mathcal{S} \\ & f_i(W) \preceq 0, \qquad i = 1, \cdots, n_c. \end{array} \quad (12)$$

In $\mathcal{H}_\infty$ analysis, we replace $\mathcal{S}$ by $\mathcal{C}$ which is an SDP representable set. For (12), we can also apply the same

procedure. The following proposition can be seen as a generalization of the Proposition 3.

Proposition 6: *Let $f_i : \mathbb{H} \to \mathbb{H}$, $i = 1 \cdots, n_c$, be the linear function. Define the set $\mathcal{F} = \{V = \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \in \mathbb{H} : f_i(W) \preceq 0, i = 1, \cdots, n_c\}$. Then $\mathbf{cl}(\mathcal{S} \cap \mathcal{F}) = \mathcal{C} \cap \mathcal{F}$.*

Proof: Since $\mathbf{cl}\,\mathcal{S} = \mathcal{C}$, and $\mathbf{cl}\,\mathcal{F} = \mathcal{F}$, we have $\mathbf{cl}\,(\mathcal{S} \cap \mathcal{F}) \subset \mathcal{C} \cap \mathcal{F}$. Now suppose $\begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \in \mathcal{C} \cap \mathcal{F}$. From Lemma 1, for any $\epsilon > 0$, there exists $\mathbf{w} \in l_2$ such that $\Lambda(\mathbf{w}) = W$, and $\|V - \Lambda(\mathbf{Mw}, \mathbf{w})\|_F < \epsilon$. Since $\Lambda(\mathbf{Mw}, \mathbf{w}) \in \mathcal{S} \cap \mathcal{F}$, we can conclude that $V \in \mathbf{cl}(\mathcal{S} \cap \mathcal{F})$. ■

Therefore, the optimization (12) is equivalent to the following SDP.

$$\begin{aligned} \underset{X,R,W}{\text{maximize}} \quad & \mathbf{Tr}\left(\begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix}\right) \\ \text{subject to} \quad & X = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \begin{bmatrix} A^* \\ B^* \end{bmatrix} \\ & \begin{bmatrix} X & R \\ R^* & W \end{bmatrix} \succeq 0, \\ & f_i(W) \preceq 0, \qquad i = 1, \cdots, n_c. \end{aligned} \tag{13}$$

Notice that as in the $\mathcal{H}_\infty$ analysis case, once the optimal solution of (13) is obtained, we can explicitly construct the worst case disturbance $\mathbf{w}$ (approximately) achieves the maximum value, and this is not available in [13].

Moreover, it is easy to derive the dual program using SDP duality and the KYP lemma like result can be obtained for a given disturbance set. For example, applying the Theorem 1 to the Example 4, the square $\mathcal{H}_\infty$ analysis, the optimal value of (13) is less than 1 if and only if there exists P, Y such that

$$\begin{aligned} Y &\succeq 0 \\ \mathbf{Tr}\,(Y) &< 1 \\ \begin{bmatrix} A^*PA - P & A^*PB \\ B^*PA & B^*PB \end{bmatrix} + \begin{bmatrix} C^*C & C^*D \\ D^*C & D^*D - Y \end{bmatrix} &\prec 0, \end{aligned}$$

where Y is a diagonal matrix. Notice that this result can also be found in [13], but the proof in here is significantly simplified.

Some of these results may be obtained obtained through the $\mathcal{S}$-procedure, but proving the $\mathcal{S}$-procedure is lossless is not a trivial task [15]. However, our approach shows that the $\mathcal{S}$-procedure can be viewed as a SDP relaxation of (12). This idea may trace back to [5], but the direct, primal formulation of the problem firstly appears here to the best of our knowledge.

V. Robust stability analysis

The robust stability analysis investigates the stability of the feedback interconnection between the nominal plant $\mathbf{G}$, a bounded operator from l_2 to l_2, and the uncertain operator Δ, a bounded operator from l_2 to l_2, which belongs to a set $\mathbf{\Delta}$. The plant $\mathbf{G}$ is said to be robustly stable with respect to $\mathbf{\Delta}$ if

$$\mathbf{I} - \Delta\mathbf{G},$$

is non-singular for all $\Delta \in \mathbf{\Delta}$. See [1] and references therein.

In this section, we investigate the possible connection between the robust stability analysis and our key Lemmas, Lemma 1 and 2.

To begin with, we assume that the uncertainty set $\mathbf{\Delta}$ admits the equivalent input-output characterization [16], [17], that is, there exists a set $\mathcal{R}_a$ such that

$$\mathcal{R}_a := \{(\mathbf{z}, \mathbf{w}) : \text{There exists } \Delta \in \mathbf{\Delta} \text{ such that } \mathbf{w} = \Delta\mathbf{z}\}.$$

Moreover, we assume that $\mathcal{R}_a$ can be completely characterized by a linear map, *i.e.*, there exists $f : \mathbb{H} \to \mathbb{H}$ such that $(\mathbf{z}, \mathbf{w}) \in \mathcal{R}_a$ if and only if

$$f(\Lambda(\mathbf{z}, \mathbf{w})) \succeq 0.$$

Let $\Lambda(\mathbf{z}, \mathbf{w}) = \begin{bmatrix} Z & R \\ R^* & W \end{bmatrix}$, then the following examples show that how f can be used to describe $\mathcal{R}_a$.

Example 7 (Full block complex LTV): *Suppose $\mathbf{\Delta} = \{\Delta : \|\Delta\| \leq 1\}$. Then $(\mathbf{z}, \mathbf{w}) \in \mathcal{R}_a$ if and only if $\|\mathbf{w}\|_2 \leq \|\mathbf{z}\|_2$. Therefore, $f(\Lambda(\mathbf{z}, \mathbf{w})) = \mathbf{Tr}\,(Z) - \mathbf{Tr}\,(W)$.*

Example 8 (Two block complex LTV, [18]): *Suppose $\mathbf{\Delta} = \{\Delta : \Delta = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}, \|\Delta_i\| \leq 1\}$. Then $(\mathbf{z}, \mathbf{w}) \in \mathcal{R}_a$ if and only if $\|\mathbf{w}_1\|_2 \leq \|\mathbf{z}_1\|_2$, and $\|\mathbf{w}_2\|_2 \leq \|\mathbf{z}_2\|_2$, where $\mathbf{z}_i, \mathbf{w}_i$ are appropriately partitioned according to the size of Δ_i. In this case, $f(\Lambda(\mathbf{z}, \mathbf{w})) = \begin{bmatrix} \mathbf{Tr}\,(Z_1) - \mathbf{Tr}\,(W_1) & 0 \\ 0 & \mathbf{Tr}\,(Z_2) - \mathbf{Tr}\,(W_2) \end{bmatrix}$.*

Example 9 (Scalar block complex LTV, [19]): *Suppose $\mathbf{\Delta} = \{\Delta : \Delta = \delta\mathbf{I}, \|\delta\| \leq 1\}$. Then $(\mathbf{z}, \mathbf{w}) \in \mathcal{R}_a$ if and only if $\Lambda(\mathbf{w}) \preceq \Lambda(\mathbf{z})$. Therefore, $f(\Lambda(\mathbf{z}, \mathbf{w})) = Z - W$.*

Example 10 (Integral quadratic constraints, [20]): *Suppose $(\mathbf{z}, \mathbf{w}) \in \mathcal{R}_a$ if and only if*

$$\sum_{k=0}^{\infty} \begin{bmatrix} z_k \\ w_k \end{bmatrix}^* \Pi \begin{bmatrix} z_k \\ w_k \end{bmatrix} \geq 0.$$

This is equivalent to

$$\mathbf{Tr}\left(\Pi \begin{bmatrix} Z & R \\ R^* & W \end{bmatrix}\right) \geq 0.$$

Notice that using this set description of $\mathcal{R}_a$, we have

$$\begin{aligned} & (\mathbf{I} - \Delta\mathbf{G})\mathbf{w} = 0 \\ \Leftrightarrow \quad & \mathbf{w} = \Delta\mathbf{G}\mathbf{w} \\ \Leftrightarrow \quad & (\mathbf{G}\mathbf{w}, \mathbf{w}) \in \mathcal{R}_a, \end{aligned}$$

This means if there exists $\|\mathbf{w}\| = 1$ such that $(\mathbf{G}\mathbf{w}, \mathbf{w}) \in \mathcal{R}_a$, then $\mathbf{G}$ is not robustly stable. Therefore it is natural to consider the following set of values generated by $\mathcal{G}$:

$$\mathcal{G} := \{f(\Lambda(\mathbf{G}\mathbf{w}, \mathbf{w})) : \|\mathbf{w}\| = 1\},$$

then we can easily see that $\mathcal{G}\cap\mathbb{H}_+\neq\varnothing$ is the equivalent condition for the existence of $\|\mathbf{w}\|=1$, such that$(\mathbf{G}\mathbf{w},\mathbf{w})\in\mathcal{R}_a$. In other words, $\mathcal{G}\cap\mathbb{H}_+\neq\varnothing$ then the system $\mathbf{G}$ cannot be robustly stable.

Therefore we can change the robust stability question to a set relationship question, but characterizing $\mathcal{G}$ may not be trivial. However, if $\mathbf{G}$ has a state-space form such that $\mathbf{x}=\mathbf{M}\mathbf{w}$, and $\mathbf{z}=C\mathbf{x}+D\mathbf{w}$, then we have the following proposition which is a direct consequence of our main result.

Proposition 7: *Suppose* $\mathbf{G}(e^{j\theta})=C(e^{j\theta}I-A)^{-1}B+D$. *Then the closure of* $\mathcal{G}$ *is given by*

$$\mathbf{cl}\,\mathcal{G}=\left\{f\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\begin{bmatrix}X & R\\ R^* & W\end{bmatrix}\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\right):\right.$$
$$\left.\mathbf{Tr}\,(W)=1,\begin{bmatrix}X & R\\ R^* & W\end{bmatrix}\in\mathcal{C}\right\}$$

Proof: Let $\mathbf{M}=(e^{j\theta}I-A)^{-1}B$. Then, $\mathbf{z}=\mathbf{G}\mathbf{w}=C\mathbf{M}\mathbf{w}+D\mathbf{w}$. Now

$$\Lambda(\mathbf{G}\mathbf{w},\mathbf{w})=\Lambda(C\mathbf{M}\mathbf{w}+D\mathbf{w},\mathbf{w})=\Lambda\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\begin{bmatrix}\mathbf{M}\mathbf{w}\\ \mathbf{w}\end{bmatrix}\right)$$
$$=\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\Lambda(\mathbf{M}\mathbf{w},\mathbf{w})\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*.$$

This shows

$$\mathcal{G}=\left\{f\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\begin{bmatrix}X & R\\ R^* & W\end{bmatrix}\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\right):\right.$$
$$\left.\mathbf{Tr}\,(W)=1,\begin{bmatrix}X & R\\ R^* & W\end{bmatrix}\in\mathcal{S}\right\}$$

Recall that $\mathbf{cl}\,\mathcal{S}=\mathcal{C}$. Using the continuity of f, we can conclude the proof. ∎

Since $\mathcal{C}$ is an SDP representable cone, the above characterization is much easier to handle compared to $\mathcal{G}$. In fact, using the theorem of alternatives, we can obtain the very interesting LMI characterization of $\mathbf{cl}\,\mathcal{G}\cap\mathbb{H}_+\neq\varnothing$.

Theorem 4 (ALT5a): *Exactly one of the following is true.*

(i) $\exists X\in\mathbb{H}$ *such that* $\mathcal{A}(X)\succeq 0$, $\mathcal{A}(X)\neq 0$, *and* $\mathcal{B}(X)=0$.

(ii) $\exists Z\in\mathbb{H}_{++}$, $W\in\mathbb{H}$, *such that* $\mathcal{A}^*(Z)+\mathcal{B}^*(W)=0$.

Proposition 8: *Exactly one of the following is true.*

(i) $\mathbf{cl}\,\mathcal{G}\cap\mathbb{H}_+=\varnothing$.

(ii) There exists $P\in\mathbb{H}$, $Y\succ 0$ *such that*

$$\begin{bmatrix}A^*PA-P & A^*PB\\ B^*PA & B^*PB\end{bmatrix}+\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*f^*(Y)\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\prec 0 \tag{14}$$

Proof: Let

$$\mathcal{A}(V)=\begin{bmatrix}V & \\ & f\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}V\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\right)\end{bmatrix}$$
$$\mathcal{B}(V)=\begin{bmatrix}A & B\end{bmatrix}V\begin{bmatrix}A & B\end{bmatrix}^*-\begin{bmatrix}I & 0\end{bmatrix}V\begin{bmatrix}I & 0\end{bmatrix}^*.$$

Then $\mathbf{cl}\,\mathcal{G}\cap\mathbb{H}_+=\varnothing$ if and only if there exists V such that $\mathcal{A}(V)\succeq 0$, but $\mathcal{A}(V)\neq 0$, and $\mathcal{B}(V)=0$. We can resale V, if necessary, to satisfy the trace condition. Notice that the adjoint of the right bottom block of $\mathcal{A}(V)$ is given by

$$\langle Y,(\mathcal{A}(V))_{22}\rangle=\mathbf{Tr}\left(Yf\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}V\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\right)\right)$$
$$=\mathbf{Tr}\left(f^*(Y)\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}V\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\right)$$
$$=\mathbf{Tr}\left(\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*f^*(Y)\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}V\right).$$

Using this fact with Theorem 4, we can conclude the proof. ∎

The above Proposition is indeed very interesting. It presents the equivalent LMI characterization of separating two sets. Let us apply the above result to the previous examples of Δ.

For Example 7, $f(V)=\mathbf{Tr}\left(\begin{bmatrix}I & 0\\ 0 & -I\end{bmatrix}V\right)$, and $f^*(y)=\begin{bmatrix}yI & 0\\ 0 & -yI\end{bmatrix}$, where the domain of f^* is $\mathbb{H}^1$. In this case, condition (14) becomes

$$\begin{bmatrix}A & B\\ 0 & I\end{bmatrix}^*\begin{bmatrix}P & 0\\ 0 & -P\end{bmatrix}\begin{bmatrix}A & B\\ 0 & I\end{bmatrix}$$
$$+\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\begin{bmatrix}yI & 0\\ 0 & -yI\end{bmatrix}\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\prec 0,$$

and since $y>0$, by multiplying $1/y$ to both sides, we recover the KYP lemma for $\|\mathbf{G}\|_\infty<1$.

For Example 9, $f(V)=\begin{bmatrix}I & 0\end{bmatrix}V\begin{bmatrix}I\\ 0\end{bmatrix}-\begin{bmatrix}0 & I\end{bmatrix}V\begin{bmatrix}0\\ I\end{bmatrix}$, and $f^*(Y)=\begin{bmatrix}Y & 0\\ 0 & -Y\end{bmatrix}$, where the domain of f^* is $\mathbb{H}^m$. In this case, condition (14) becomes

$$\begin{bmatrix}A & B\\ 0 & I\end{bmatrix}^*\begin{bmatrix}P & 0\\ 0 & -P\end{bmatrix}\begin{bmatrix}A & B\\ 0 & I\end{bmatrix}$$
$$+\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\begin{bmatrix}Y & 0\\ 0 & -Y\end{bmatrix}\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\prec 0,$$

Since $Y\succ 0$, by left and right multiplying $Y^{-1/2}$ to the above expression, we can recover $\|Y^{-1/2}\mathbf{G}Y^{1/2}\|_\infty<1$, which is a scaled small gain test.

Finally, for Example 10, we have $f(V)=\mathbf{Tr}\,(\Pi V)$, and $f^*(y)=y\Pi$, where the domain is $\mathbb{H}^1$. In this case, condition (14) becomes

$$\begin{bmatrix}A^*PA-P & A^*PB\\ B^*PA & B^*PB\end{bmatrix}+\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^*\Pi\begin{bmatrix}C & D\\ 0 & I\end{bmatrix}\prec 0.$$

For block diagonal structure, such as Example 8 we can also apply our approach to recover the block diagonal small gain test as in [18].

From robust control theory, we know that the scaled small gain test provides a sufficient and necessary test for robust stability in certain cases [1]. In fact, all the examples we provide fall in to those classes. Notice that we have shown that the scaled small gain test is a necessary and sufficient condition for $\mathbf{cl}\,\mathcal{G}\cap\mathbb{H}_+=\varnothing$. Therefore, we obtain the following chain of equivalent statements when the scaled

small gain test becomes the equivalent condition for the robust stability of $\mathbf{G}$.

$$\begin{aligned} \mathbf{cl}\,\mathcal{G} \cap \mathbb{H}_+ = \varnothing &\overset{(a)}{\Leftrightarrow} \|\Theta^{-1}\mathbf{G}\Theta\|_\infty < 1 \\ \overset{(b)}{\Leftrightarrow} &\text{ Robust stability of } \mathbf{G} \end{aligned}$$

However, proving (b) from the scaled small gain test is not a trivial task whereas (a) is from a standard machinery. We strongly believe that there exists a direct proof between ($\mathbf{cl}\,\mathcal{G} \cap \mathbb{H}_+ = \varnothing$)⇔(Robust stability of $\mathbf{G}$) in the style of [21] without relying on the complicated argument, and this is currently under investigation.

Another important observation is that if $\mathbf{cl}\,\mathcal{G} \cap \mathbb{H}_+ \neq \varnothing$, then we can use SDP to find such a V. Then using Lemma 1, we can find a pair $(\mathbf{w}, \mathbf{z})$ that approximately satisfies

$$\begin{aligned} \mathbf{z} &= \mathbf{G}\mathbf{w} \\ \mathbf{w} &= \Delta\mathbf{z}. \end{aligned}$$

Notice that this pair $(\mathbf{w}, \mathbf{z})$ disproves the robust stability.

In conclusion, we have shown how our main result can be used to derive the scaled small gain test without using the commutant, Θ. Moreover, we show that the exact implication of the scaled small gain test, $\mathbf{cl}\,\mathcal{G} \cap \mathbb{H}_+ = \varnothing$, using our results. This suggests that there may exist deeper connection between robust stability and the set relationship $\mathbf{cl}\,\mathcal{G}\cap\mathbb{H}_+ = \varnothing$, and we are currently investigating their exact relationship.

VI. Conclusion

In this paper, we propose an alternative, reverse direction of theoretical development for robust control theory. Based on our Lemma 1, an SDP representation of a set of gramians, we show that the robust analysis and stability question can be directly formulated as a primal optimization. Moreover, we show that the well-known results in robust control theory can be obtained via SDP duality, an arguably simple machinery to prove many interesting results, and this shows that our approach is a primal formulation of robustness analysis. Therefore, we believe that our paper provides an alternative ”primal-dual” picture in robust control theory, and this new development not only opens up the new research direction, but also enhances pedagogy.

VII. Appendix: Proof of the main result

This section we present the main proof of our proposition 1, with technical lemmas that need to prove our result.

A. Preliminaries from linear algebra

The following results from the linear algebra are used to prove the main result.

Proposition 9: *For $x, y \in \mathbb{C}^n$, $\|xy^*\|_F = \|x\|_2\|y\|_2$.*

Proof: $\|xy^*\|_F^2 = \mathbf{Tr}\,((xy^*)^*(yx^*)) = \|x\|_2^2\|y\|_2^2$. ■

Theorem 5 (Gelfand, 1941): *For any matrix norm $\|\cdot\|$,*

$$\lim_{n\to\infty} \|A^n\|^{1/n} = \rho(A).$$

Proposition 10: *Suppose $\rho(A) < 1$. Then, for any matrix norm $\|\cdot\|$, $\|A^k\| \in l_1$.*

Proof: Let $\epsilon = (1 - \rho(A))/2 > 0$. Then from the Theorem 5, there exists $N \in \mathbb{N}$ such that

$$\|A^k\| < (\rho(A) + \epsilon)^k,$$

for all $k \geq N$.

$$\begin{aligned} \sum_{k=0}^{\infty} \|A^k\| &= \sum_{k=0}^{N} \|A^k\| + \sum_{k=N+1}^{\infty} \|A^k\| \\ &< \sum_{k=0}^{N} \|A^k\| + \sum_{k=N+1}^{\infty} (\rho(A) + \epsilon)^k. \end{aligned}$$

Since $\rho(A) + \epsilon < 1$, the second term is finite. Therefore, $\|A^k\| \in l_1$. ■

Lemma 3 (Rantzer, 1996): *Let F, G complex matrices with same dimension. Then $FF^* = GG^*$ if and only if there exists a unitary matrix $U \in \mathbb{C}^{k\times k}$ such that $F = GU$.*

Proof: See [9]. ■

The following result about a convex cone $\mathcal{C}$ is a direct consequence of the above lemma, and the results states that extreme points of $\mathcal{C}$ are rank one matrices.

Proposition 11 (Rank one decomposition): *For all $V \in \mathcal{C}$, there exists a set of matrices $V_1, \cdots, V_{n+m} \in \mathcal{C}$ such that $V = \sum_{k=1}^{n+m} V_k$, and $\mathbf{Rank}\,(V_k) \leq 1$ for all $k = 1, \cdots, n+m$.*

Proof: See [8]. ■

B. Technical lemmas for the main result

In this section, we derive technical results to prove the main result of this paper. The basic idea behind the following results is to bound the error terms arise in the proof of our main result.

Proposition 12: *Suppose $\mathbf{w}$ has finite number of non-zero entries. For any $\varepsilon > 0$, there exists $N \in \mathbb{N}$, such that*

$$\|\Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) - \Lambda_n(\mathbf{M}\mathbf{w}, \mathbf{w})\|_F < \varepsilon,$$

for all $n \geq N$.

Proof: Let $\mathbf{x} = \mathbf{M}\mathbf{w}$, and consider $T \in \mathbb{N}$ such that $w_k = 0$, for all $k \geq T$. For $N \geq T$, we have

$$\begin{aligned} &\|\Lambda(\mathbf{M}\mathbf{w}, \mathbf{w}) - \Lambda_N(\mathbf{M}\mathbf{w}, \mathbf{w})\|_F \\ = \; &\left\| \sum_{k=N}^{\infty} A^k x_T (A^k x_T)^* \right\|_F \leq \sum_{k=N}^{\infty} \|A^k x_T\|_2^2 \\ = \; &\|A^N x_T\|_2^2 \sum_{k=0}^{\infty} \|A^k\|_2^2 \end{aligned}$$

Since $\|A^k\|_2 \in l_1 \subset l_2$, the infinite sum is finite. In addition, from the Theorem 5, $\|A^N x_T\|_F \to 0$ as $N \to \infty$. Therefore by choosing N sufficiently large, we can conclude the proof. ■

Proposition 13: *Let $\mathbf{x}, \mathbf{w} \in l_2$, and $\mathbf{y} \in l_2$ such that $y_k = A^k y_0$ for all k. Then there exists a constant C such that*

$$\begin{aligned} &\|\Lambda(\mathbf{x}+\mathbf{y},\mathbf{w}) - \Lambda(\mathbf{x},\mathbf{w})\| \\ \leq\ & C\max\{(\|\mathbf{x}\|_\infty + \|\mathbf{w}\|_\infty)\|y_0\|_2, \|y_0\|_2^2\}. \end{aligned}$$

Proof: Notice that,

$$\begin{aligned} &\|\Lambda(\mathbf{x}+\mathbf{y},\mathbf{w}) - \Lambda(\mathbf{x},\mathbf{w})\|_F \\ =\ & \left\| \sum_{k=0}^{\infty} \begin{bmatrix} x_k + y_k \\ w_k \end{bmatrix} \begin{bmatrix} x_k + y_k \\ w_k \end{bmatrix}^* - \begin{bmatrix} x_k \\ w_k \end{bmatrix} \begin{bmatrix} x_k \\ w_k \end{bmatrix}^* \right\|_F \\ =\ & \left\| \sum_{k=0}^{\infty} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \begin{bmatrix} x_k \\ w_k \end{bmatrix}^* + \begin{bmatrix} x_k \\ w_k \end{bmatrix} \begin{bmatrix} y_k \\ 0 \end{bmatrix}^* + \begin{bmatrix} y_k \\ 0 \end{bmatrix} \begin{bmatrix} y_k \\ 0 \end{bmatrix}^* \right\|_F \\ \leq\ & \sum_{k=0}^{\infty} 2 \left\| \begin{bmatrix} y_k \\ 0 \end{bmatrix} \begin{bmatrix} x_k \\ w_k \end{bmatrix}^* \right\|_F + \left\| \begin{bmatrix} y_k \\ 0 \end{bmatrix} \begin{bmatrix} y_k \\ 0 \end{bmatrix}^* \right\|_F \\ =\ & \sum_{k=0}^{\infty} 2\|y_k\|_2 \sqrt{\|x_k\|_2^2 + \|w_k\|_2^2} + \|y_k\|_2^2. \end{aligned}$$

Since $\mathbf{x}, \mathbf{w} \in l_2 \subset l_2$, we have

$$\sqrt{\|x_k\|_2^2 + \|w_k\|_2^2} \leq \|\mathbf{x}\|_\infty + \|\mathbf{w}\|_\infty,$$

for all k. Moreover, since $y_k = A^k y_0$, we have

$$\begin{aligned} &\|\Lambda(\mathbf{x}+\mathbf{y},\mathbf{w}) - \Lambda(\mathbf{x},\mathbf{w})\|_F \\ \leq\ & \sum_{k=0}^{\infty} 2(\|\mathbf{x}\|_\infty + \|\mathbf{w}\|_\infty)\|A^k y_0\|_2 + \|A^k y_0\|_2^2 \\ \leq\ & \sum_{k=0}^{\infty} 2(\|\mathbf{x}\|_\infty + \|\mathbf{w}\|_\infty)\|A^k\|_2\|y_0\|_2 + \|A^k\|_2^2\|y_0\|_2^2 \\ \leq\ & C\max\{(\|\mathbf{x}\|_\infty + \|\mathbf{w}\|_\infty)\|y_0\|_2, \|y_0\|_2^2\} \end{aligned}$$

where $C = \left(\sum_{k=0}^{\infty} 2\|A^k\|_2 + \|A^k\|_2^2\right)$. Since $\|A^k\|_2 \in l_1 \subset l_2$, $C < \infty$, and this concludes the proof. ∎

C. Main result

Now we are ready to prove main results of this paper. The main idea of the proof is as follows.

Any rank one matrix in $\mathcal{C}$ can be generated by a sinusoid $\mathbf{w}$, but a sinusoid is not in l_2. Therefore we find a signal in l_2 which approximates this sinusoid. This is not surprising, because in $\mathcal{H}_\infty$ analysis [22], the so called worst case signal is sinusoid which is not in l_2, so one has to approximate this sinusoid using l_2 and the supremum is not achieved. More fundamental reason for this is due to non-compactness of an unit sphere in l_2, but we will not elaborate this point.

For a full rank matrix in $\mathcal{C}$, we firstly decompose this matrix to rank one matrices using the Lemma 11, then approximate each rank one matrices by a signal in l_2. Finally, we pad them together to approximate a full rank matrix as in [18].

Proposition 14: *Suppose $V \in \mathcal{C}$, and $\mathbf{Rank}\,(V) \leq 1$. Then for all $\varepsilon > 0$, there exists $\mathbf{w}$ with a finite number of non-zero entries such that*

$$\|\Lambda(\mathbf{M}\mathbf{w},\mathbf{w}) - V\|_F < \varepsilon \tag{15}$$

$$\Lambda(\mathbf{w}) = \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix} V \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix}^* \tag{16}$$

Proof: Suppose $\mathbf{Rank}\,(V) = 0$. Then $V = 0 \in \mathcal{C}$, and $\mathbf{w} = 0$ satisfies (15) and (16).

Now suppose $\mathbf{Rank}\,(V) = 1$. We will construct $\mathbf{w}$ which satisfies (15) and (16). Since $\mathbf{Rank}\,(V) = 1$, there exists $x_s \in \mathbb{C}^n, w_s \in \mathbb{C}^m$ such that $V = \begin{bmatrix} x_s \\ w_s \end{bmatrix} \begin{bmatrix} x_s \\ w_s \end{bmatrix}^*$. Moreover, by defining $f = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x_s \\ w_s \end{bmatrix}$, $g = \begin{bmatrix} I_n & 0_{n,m} \end{bmatrix} \begin{bmatrix} x_s \\ w_s \end{bmatrix}$, we can easily see that $ff^* = gg^*$. Therefore from the Lemma 3, there exists θ such that $e^{j\theta}x_s = Ax_s + Bw_s$.

Now, for a given $N \in \mathbb{N}$, define $\mathbf{w}$

$$w_k = \begin{cases} \frac{1}{\sqrt{N}} e^{j\theta k} w_s & \text{if } 0 \leq k < N \\ 0 & \text{if } N \leq k. \end{cases}$$

It is easy to see that $\Lambda(\mathbf{w}) = \sum_{k=0}^{\infty} w_k w_k^* = \sum_{k=0}^{N-1} w_k w_k^* = w_s w_s^*$. and therefore $\mathbf{w}$ satisfies (16). In order to obtain $\mathbf{x} = \mathbf{M}\mathbf{w}$, let us define the following signal $\mathbf{s}, \mathbf{t}$

$$\begin{aligned} s_k &= \begin{cases} \frac{1}{\sqrt{N}} e^{j\theta_k} x_s & \text{if } 0 \leq k < N \\ \frac{1}{\sqrt{N}} A^{k-N} e^{j\theta N} x_s & \text{if } N \leq k, \end{cases} \\ t_k &= -\frac{1}{\sqrt{N}} A^k x_s \end{aligned}$$

then $\mathbf{x} = \mathbf{s} + \mathbf{t}$. Notice that

$$\Lambda_N(\mathbf{s},\mathbf{w}) = \sum_{k=0}^{N-1} \begin{bmatrix} s_k \\ w_k \end{bmatrix} \begin{bmatrix} s_k \\ w_k \end{bmatrix}^* = \frac{1}{N} \sum_{k=0}^{N-1} \begin{bmatrix} x_s \\ w_s \end{bmatrix} \begin{bmatrix} x_s \\ w_s \end{bmatrix}^* = V$$

and this shows

$$\begin{aligned} &\|\Lambda(\mathbf{s},\mathbf{w}) - V\|_F = \left\| \sum_{k=N}^{\infty} \begin{bmatrix} s_k \\ w_k \end{bmatrix} \begin{bmatrix} s_k \\ w_k \end{bmatrix}^* \right\|_F \\ \leq\ & \frac{1}{N} \sum_{k=0}^{\infty} \left\| \begin{bmatrix} A^k x_s \\ 0 \end{bmatrix} \begin{bmatrix} A^k x_s \\ 0 \end{bmatrix}^* \right\|_F \leq \frac{1}{N} \sum_{k=0}^{\infty} \|A^k x_s\|_2^2 \\ \leq\ & \underbrace{\sum_{k=0}^{\infty} \|A^k\|_2^2}_{C_1} \frac{\|x_s\|_2^2}{N}. \end{aligned}$$

Notice that $C_1 < \infty$ because of the Proposition 10.

Now using the triangle inequality and the Proposition 13, we obtain

$$\begin{aligned} &\|\Lambda(\mathbf{x},\mathbf{w}) - V\|_F = \|\Lambda(\mathbf{s}+\mathbf{t},\mathbf{w}) - V\|_F \\ \leq\ & \|\Lambda(\mathbf{s},\mathbf{w}) - V\|_F + \|\Lambda(\mathbf{s}+\mathbf{t},\mathbf{w}) - \Lambda(\mathbf{s},\mathbf{w})\|_F \\ \leq\ & C_1 \frac{\|x_s\|_2^2}{N} + C\max\{(\|\mathbf{s}\|_\infty + \|\mathbf{w}\|_\infty)\|t_0\|_2, \|t_0\|_2^2\} \end{aligned}$$

Since $\|\mathbf{s}\|_\infty = \frac{1}{\sqrt{N}}\|x_s\|$, $\|\mathbf{w}\|_\infty = \frac{1}{\sqrt{N}}\|w_s\|$, $\|t_0\|_2 = \frac{1}{\sqrt{N}}\|x_s\|$, we have $\max\{(\|\mathbf{s}\|_\infty + \|\mathbf{w}\|_\infty)\|t_0\|_2, \|t_0\|_2^2\} = \frac{1}{N}\|x_s\|_2(\|x_s\|_2 + \|w_s\|_2)$. By combining all these bounds,

we can conclude that there exists a constant C_3 only depends on A, x_s, w_s such that

$$\|\Lambda(\mathbf{x},\mathbf{w}) - V\|_F \leq \frac{C_3}{N},$$

therefore by taking sufficiently large N, we can make $\mathbf{w}$ satisfy (15), and clearly $\mathbf{w}$ has N number of non-zero entries. ∎

Now combining all these results, we are ready to prove our main result, the Proposition 1. *Proof:* [Proof of the Proposition 1] From the Lemma 11, we can decompose $V = \sum_{i=1}^{n+m} V_i$ where $V_i \in \mathcal{C}$, and $\mathbf{Rank}\,(V_i) \leq 1$. Let us rearrange these terms, so that $V = \sum_{i=1}^{r} V_i$ where $\mathbf{Rank}\,(V_i) = 1$. We now use an induction on r. Suppose $r \leq 1$, then from the Proposition 14, the proof is done.

Now assume the induction hypothesis holds, that is for $\sum_{i=1}^{r-1} V_i \in \mathcal{C}$, there exists $\tilde{\mathbf{w}}$ with a finite number of non-zero entries such that

$$\left\|\Lambda(\mathbf{M}\tilde{\mathbf{w}}, \tilde{\mathbf{w}}) - \sum_{i=1}^{r-1} V_i\right\|_F < \frac{1}{4}\varepsilon,$$

$$\Lambda(\tilde{\mathbf{w}}) = \sum_{i=1}^{r-1} \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix} V_i \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix}^*$$

Similarly, for V_r, there exists $\hat{\mathbf{w}}$ with a finite number of non-zero entries such that

$$\|\Lambda(\mathbf{M}\hat{\mathbf{w}}, \hat{\mathbf{w}}) - V_r\|_F < \frac{1}{4}\varepsilon,$$

$$\Lambda(\hat{\mathbf{w}}) = \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix} V_r \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix}^*$$

satisfies with $\varepsilon/4$ and 2. In addition, since $V_r \in \mathcal{C}$ and $\mathbf{Rank}\,(V_r) = 1$, from the Proposition 14, there exists finite length $\hat{\mathbf{w}}$ satisfies 1 with $\varepsilon/4$ and 2.

Let $T \in \mathbb{N}$ such that $\tilde{w}_k = 0$ for all $k \geq T$. From the Proposition 12, we can find N_1 such that

$$\|\Lambda(\mathbf{M}\tilde{\mathbf{w}}, \tilde{\mathbf{w}}) - \Lambda_{n+T}(\mathbf{M}\tilde{\mathbf{w}}, \tilde{\mathbf{w}})\| < \frac{1}{4}\epsilon,$$

for all $n \geq N_1$.

Consider the following signal $\mathbf{w}$

$$w_k = \begin{cases} \tilde{w}_k & \text{if } 0 \leq k < N+T \\ \hat{w}_{k-N-T} & \text{if } N+T \leq k, \end{cases}$$

where $N \geq N_1$. Clearly, $\mathbf{w}$ has a finite number of non-zero entries, and $\Lambda(\mathbf{w}) = \Lambda(\tilde{\mathbf{w}}) + \Lambda(\hat{\mathbf{w}})$, which shows

$$\Lambda(\mathbf{w}) = \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix} V \begin{bmatrix} 0_{m,n} & I_m \end{bmatrix}^*.$$

Now, let $\tilde{\mathbf{x}} = \mathbf{M}\tilde{\mathbf{w}}$, $\hat{\mathbf{x}} = \mathbf{M}\hat{\mathbf{w}}$, and $\mathbf{x} = \mathbf{M}\mathbf{w}$. Then,

$$x_k = \begin{cases} \tilde{x}_k & \text{if } 0 \leq k < T \\ A^{k-T}\tilde{x}_T & \text{if } T \leq k < N+T \\ \hat{x}_{k-N-T} + A^{k-T}\tilde{x}_T & \text{if } N+T \leq k. \end{cases}$$

Notice that

$$\begin{aligned} \Lambda(\mathbf{x},\mathbf{w}) &= \Lambda_{N+T}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) \\ &\quad + \sum_{k=0}^{\infty} \begin{bmatrix} \hat{x}_k + A^k A^N \tilde{x}_T \\ \hat{w}_k \end{bmatrix} \begin{bmatrix} \hat{x}_k + A^k A^N \tilde{x}_T \\ \hat{w}_k \end{bmatrix}^* \\ &= \Lambda_{N+T}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) + \Lambda(\hat{\mathbf{x}} + \mathbf{y}, \hat{\mathbf{w}}), \end{aligned}$$

where $y_k = A^k A^N \tilde{x}_T$.

Therefore,

$$\begin{aligned} &\|\Lambda(\mathbf{x},\mathbf{w}) - V\|_F \\ \leq\ & \left\|\Lambda_{N+T}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) - \sum_{i=1}^{r-1} V_i\right\|_F + \|\Lambda(\hat{\mathbf{x}} + \mathbf{y}, \hat{\mathbf{w}}) - V_r\|_F \\ \leq\ & \|\Lambda_{N+T}(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) - \Lambda(\tilde{\mathbf{x}}, \tilde{\mathbf{w}})\|_F + \left\|\Lambda(\tilde{\mathbf{x}}, \tilde{\mathbf{w}}) - \sum_{i=1}^{r-1} V_i\right\|_F \\ & + \|\Lambda(\hat{\mathbf{x}} + \mathbf{y}, \hat{\mathbf{w}}) - \Lambda(\hat{\mathbf{x}}, \hat{\mathbf{w}})\|_F + \|\Lambda(\hat{\mathbf{x}}, \hat{\mathbf{w}}) - V_r\|_F \\ \leq\ & \frac{3}{4}\epsilon + C\max\{(\|\hat{\mathbf{x}}\|_\infty + \|\hat{\mathbf{w}}\|_\infty)\|y_0\|_2, \|y_0\|_2^2\}. \end{aligned}$$

Notice that $\hat{\mathbf{x}}, \hat{\mathbf{w}}$ are given signals in l_2, and does not depend on our choice N. However, since $\|y_0\|_2 = \|A^N\|\|\tilde{x}\|_T$ can be made arbitrarily small by taking $N \to \infty$, and this concludes the proof. ∎

References

[1] G. E. Dullerud and F. Paganini, *A course in robust control theory*. Springer New York, 2000.

[2] S. Boyd, L. E. Ghaoui, E. Feron, V. Balakrishnan, and V. Yakubovich, "Linear matrix inequalities in system and control theory," *SIAM Review*, vol. 37, no. 3, pp. 479–480, 1995.

[3] V. A. Yakubovich, "S-procedure in nonlinear control theory," *Vestnik Leningrad University*, vol. 1, pp. 62–77, 1971.

[4] V. Balakrishnan and L. Vandenberghe, "Semidefinite programming duality and linear time-invariant systems," *Automatic Control, IEEE Transactions on*, vol. 48, no. 1, pp. 30–41, 2003.

[5] C. W. Scherer, "LMI relaxations in robust control," *European Journal of Control*, vol. 12, no. 1, pp. 3–29, 2006.

[6] Y. Ebihara, Y. Onishi, and T. Hagiwara, "Robust performance analysis of uncertain lti systems: Dual lmi approach and verifications for exactness," *Automatic Control, IEEE Transactions on*, vol. 54, no. 5, pp. 938–951, 2009.

[7] A. Gattami and B. Bamieh, "A simple approach to H_∞ analysis," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 2424–2428.

[8] S. You and A. Gattami, "H infinity analysis revisited," *arXiv preprint arXiv:1412.6160*, 2014.

[9] A. Rantzer, "On the Kalman–Yakubovich–Popov lemma," *Systems & Control Letters*, vol. 28, no. 1, pp. 7–10, 1996.

[10] T. T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *Automatic Control, IEEE Transactions on*, vol. 47, no. 7, pp. 1056–1066, 2002.

[11] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3a MATLAB software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[12] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Society For Industrial Mathematics, 1987, vol. 2.

[13] R. D'Andrea, "H_∞ optimization with spatial constraints," in *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, vol. 4. IEEE, 1995, pp. 4327–4332.

[14] F. Paganini, "Convex methods for robust H_2 analysis of continuous-time systems," *Automatic Control, IEEE Transactions on*, vol. 44, no. 2, pp. 239–252, 1999.

[15] I. Pólik and T. Terlaky, "A survey of the S-lemma," *SIAM review*, vol. 49, no. 3, pp. 371–418, 2007.

[16] F. Paganini, R. D'Andrea, and J. Doyle, "Behavioral approach to robustness analysis," in *American Control Conference, 1994*, vol. 3. IEEE, 1994, pp. 2782–2786.

[17] R. D'Andrea, F. Paganini, and J. C. Doyle, "Uncertain behavior," 1993.

[18] J. S. Shamma, "Robust stability with time-varying structured uncertainty," *Automatic Control, IEEE Transactions on*, vol. 39, no. 4, pp. 714–724, 1994.

[19] F. Paganini-Herrera, "Sets and constraints in the analysis of uncertain systems," Ph.D. dissertation, California Institute of Technology, 1996.

[20] A. Megretski and A. Rantzer, “System analysis via integral quadratic constraints,” *Automatic Control, IEEE Transactions on*, vol. 42, no. 6, pp. 819–830, 1997.
[21] M. G. Safonov, *Stability and robustness of multivariable feedback systems*. MIT press, 1980.
[22] J. C. Doyle, B. A. Francis, and A. Tannenbaum, *Feedback control theory*. Macmillan Publishing Company New York, 1992.